Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of particulate matter.

Study Population/Test Model
Controlled Human Exposure
In general, the subjects recruited into study groups should be similarly matched for age, sex, race, anthropometric properties, and health status. In studies evaluating effects of specific subject characteristics (e.g., disease, genetic polymorphism, etc.), appropriately matched healthy controls are preferred. Relevant characteristics and health status should be reported for each experimental group. Criteria for including and excluding subjects should be clearly indicated. For the examination of populations with an underlying health condition (e.g., asthma), independent, clinical assessment of the health condition is ideal, but self-report of physician diagnosis generally is considered to be reliable for respiratory and cardiovascular disease outcomes[b]. The loss or withdrawal of recruited subjects during the course of a study should be reported. Specific rationale for excluding subject(s) from any portion of a protocol should be explained.
Animal Toxicology
Ideally, studies should report species, strain, substrain, genetic background, age, sex, and weight. Unless data indicate otherwise, all animal species and strains are considered appropriate for evaluating effects of PM exposure. It is preferred that the authors test for effects in both sexes and multiple lifestages, and report the result for each group separately. All animals used in a study should be accounted for, and rationale for exclusion of animals or data should be specified.
Epidemiology
There is greater confidence in results for study populations that are recruited from and representative of the target population. Studies with high participation and low drop-out over time that is not dependent on exposure or health status are considered to have low potential for selection bias. Clearly specified criteria for including and excluding subjects can aid assessment of selection bias. For populations with an underlying health condition, independent, clinical assessment of the health condition is valuable, but self-report of physician diagnosis generally is considered to be reliable for respiratory and cardiovascular diseases[b]. Comparisons of groups with and without an underlying health condition are more informative if groups are from the same source population. Selection bias can influence results in either direction or may not affect the validity of results but rather reduce the generalizability of findings to the target population.
Pollutant
Controlled Human Exposure
Studies should: (1) include a composite measure of PM (i.e., $PM_{2.5}$, $PM_{10-2.5}$, or ultrafine particles [UFP][c]) or (2) apply some approach (e.g., particle trap or filter) to assess the effects of PM in a complex air pollution mixture (i.e., diesel exhaust, gasoline exhaust, wood smoke).
Animal Toxicology
Studies should: (1) include a composite measure of PM (i.e., $PM_{2.5}$, $PM_{10-2.5}$, or ultrafine particles [UFP][c]) or (2) apply some approach (e.g., particle trap or filter) to assess the effects of PM in a complex air pollution mixture (i.e., diesel exhaust, gasoline exhaust, wood smoke).
Epidemiology
Health effects are evaluated primarily using a composite measure of PM (i.e., $PM_{2.5}$, $PM_{10-2.5}$, or ultrafine particles [UFP][c]) from studies using ambient measurements, model predictions, or a combination of measured and modeled data. Studies of PM components must also include a composite measure of PM. Studies of source-related indicators are also evaluated where the indicator is derived using ambient PM concentrations.

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of particulate matter.

Exposure Assessment or Assignment
Controlled Human Exposure
For this assessment, the focus is on studies that utilize PM concentrations <2 mg/m^3. Studies that use higher exposure concentrations may provide information relevant to biological plausibility, dosimetry, or inter-species variation. Studies should have well-characterized pollutant concentration, temperature, and relative humidity and/or have measures in place to adequately control the exposure conditions. Preference is given to balanced crossover or parallel design studies which include control exposures (e.g., to clean filtered air). Study subjects should be randomly exposed without knowledge of the exposure condition. Method of exposure (e.g., chamber, facemask, etc.) should be specified and activity level of subjects during exposures should be well characterized.
Animal Toxicology
For this assessment, the focus is on studies that utilize PM concentrations <2 mg/m^3. Studies that use higher exposure concentrations may provide information relevant to biological plausibility, dosimetry, or inter-species variation. Studies should characterize pollutant concentration, temperature, and relative humidity and/or have measures in place to adequately control the exposure conditions. The focus is on inhalation exposure. Non-inhalation exposure experiments (i.e., intratracheal instillation [IT]) are informative for size fractions (e.g., $PM_{10-2.5}$) that cannot penetrate the airway of a study animal and may provide information relevant to biological plausibility and dosimetry. In vitro studies may be included if they provide mechanistic insight or examine similar effects as in vivo studies, but are generally not included. All studies should include exposure control groups (e.g., clean filtered air).

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of particulate matter.

Epidemiology
Of primary relevance are relationships of health effects with the ambient component of PM exposure. However, information about ambient exposure rarely is available for individual subjects; most often, inference is based on ambient concentrations. Studies that compare exposure assessment methods are considered to be particularly informative. Inference is stronger when the duration or lag of the exposure metric corresponds with the time course for physiological changes in the outcome (e.g., up to a few days for symptoms) or latency of disease (e.g., several years for cancer). Given that the spatial variability of PM composite measures varies among size fractions, with more homogeneity for $PM_{2.5}$ than either $PM_{10-2.5}$ or UFP, the need for capturing spatial contrasts is stronger for $PM_{10-2.5}$ or UFP compared with $PM_{2.5}$. Validated measurements, whether averaged across multiple monitors or assigned from the nearest or single available monitor, adequately capture temporal or spatial variation in exposure to $PM_{2.5}$ due to the high correlation between personal exposure and ambient concentration. However, for more spatially heterogeneous $PM_{10-2.5}$ and UFP, the spatial correlation between personal exposure and ambient concentrations is lower. Similarly, PM components show increased spatial variability relative to $PM_{2.5}$. In this case, validated methods that capture the extent of variability for the particular study design (temporal vs. spatial contrasts) and location carry greater weight. Inference based on central site measurements can be adequate if correlated with personal exposures, closely located to study subjects, highly correlated across monitors within a location, used in locations with well-distributed sources, or combined with time-activity information. In studies of short-term exposure, temporal variability of the exposure metric is of primary interest. For all PM size fractions, studies that incorporate time-activity data with personal or microenvironmental monitoring or modeling data may carry greater weight because residential, in-vehicle, and workplace PM exposures may differ in their temporal variability. Results for total personal and indoor PM exposure are other lines of evidence that may inform judgments about causality of PM because inference is based on an individual's microenvironmental exposures and the potential for copollutant confounding may be reduced compared to ambient exposures. Results for total personal exposure can inform understanding of the effects of ambient exposure when well correlated with ambient concentrations. For long-term exposures, methods that well represent within-community spatial variation in individual exposure may be given more weight for spatially-variable ambient $PM_{10-2.5}$ or ultrafine particles. For $PM_{2.5}$, within-community variation in exposure is less important given that $PM_{2.5}$ tends to be more homogeneous. Exposure measurement error often attenuates health effect estimates or increases the imprecision of the association (i.e., width of 95% CIs), particularly associations based on temporal variation in short-term exposure. However, exposure measurement error can bias estimates away from the null in some epidemiologic studies of long-term exposures where the PM size fraction is more spatially heterogeneous (i.e., $PM_{10-2.5}$ or UFP), depending on the locations of the monitor and sources with respect to the study population. To streamline the health effects discussion on studies that are most policy-relevant, for those health categories where the 2009 PM ISA concluded a "causal relationship" the focus is on studies with mean $PM_{2.5}$ concentrations <20 $\mu g/m^3$. However, studies that examine a previously identified uncertainty or limitation in the evidence are evaluated even if mean $PM_{2.5}$ concentrations are >20 $\mu g/m^3$.
Outcome Assessment/Evaluation
Controlled Human Exposure
Endpoints should be assessed in the same manner for control and exposure groups (e.g., time after exposure, methods, endpoint evaluator) using valid, reliable methods. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints (e.g., histopathology). For each experiment and each experimental group, including controls, precise details of all procedures carried out should be provided including how, when, and where. Time of the endpoint evaluations is a key consideration that will vary depending on endpoint evaluated. Endpoints should be assessed at time points that are appropriate for the research questions.

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of particulate matter.

Animal Toxicology
Endpoints should be assessed in the same manner for control and exposure groups (e.g., time after exposure, methods, endpoint evaluator) using valid, reliable methods. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints (e.g., histopathology). For each experiment and each experimental group, including controls, precise details of all procedures carried out should be provided including how, when, and where. Time of the endpoint evaluations is a key consideration that will vary depending on endpoint evaluated. Endpoints should be assessed at time points that are appropriate for the research questions.
Epidemiology
Inference is stronger when outcomes are assessed or reported without knowledge of exposure status. Knowledge of exposure status could produce artefactual associations. Confidence is greater when outcomes assessed by interview, self-report, clinical examination, or analysis of biological indicators are defined by consistent criteria and collected by validated, reliable methods. Independent, clinical assessment is valuable for outcomes such as lung function or incidence of disease, but report of physician diagnosis has shown good reliability[b]. When examining short-term exposures, evaluation of the evidence focuses on specific lags based on the evidence presented in individual studies. Specifically, the following hierarchy is used in the process of selecting results from individual studies to assess in the context of results across all studies for a specific health effect or outcome: • Distributed lag models; • Average of multiple days (e.g., 0–2); • If a priori lag days were used by the study authors these are the effect estimates presented; or • If a study focuses on only a series of individual lag days, expert judgment is applied to select the appropriate result to focus on considering the time course for physiologic changes for the health effect or outcome being evaluated. When health effects of long-term exposure are assessed by acute events such as symptoms or hospital admissions, inference is strengthened when results are adjusted for short-term exposure. Validated questionnaires for subjective outcomes such as symptoms are regarded to be reliable[c], particularly when collected frequently and not subject to long recall. For biological samples, the stability of the compound of interest and the sensitivity and precision of the analytical method is considered. If not based on knowledge of exposure status, errors in outcome assessment tend to bias results toward the null.
Potential Copollutant Confounding
Controlled Human Exposure
Exposure should be well characterized to evaluate independent effects of PM of various size fractions. Studies should apply some approach (e.g., particle trap or filter) to assess the effects of PM when examining exposures to complex air pollution mixtures (i.e., diesel exhaust, gasoline exhaust, wood smoke).
Animal Toxicology
Exposure should be well characterized to evaluate independent effects of PM of various size fractions. Studies should apply some approach (e.g., particle trap or filter) to assess the effects of PM when examining exposures to complex air pollution mixtures (i.e., diesel exhaust, gasoline exhaust, wood smoke).

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of particulate matter.

Epidemiology
Not accounting for potential copollutant confounding can produce artefactual associations; thus, studies that examine copollutant confounding carry greater weight. The predominant method is copollutant modeling (i.e., two-pollutant models), which is especially informative when correlations are not high. However, when correlations are high ($r > 0.7$), such as those often encountered for UFP and other traffic-related copollutants, copollutant modeling is less informative. Although the use of single-pollutant models to examine the association between PM and a health effect or outcome are informative, ideally studies should also include copollutant analyses. Copollutant confounding is evaluated on an individual study basis considering the extent of correlations observed between the copollutant and PM, and relationships observed with PM and health effects in copollutant models.
Other Potential Confounding Factors[d]
Controlled Human Exposure
Preference is given to studies utilizing experimental and control groups that are matched for individual level characteristics (e.g., race/ethnicity, sex, body weight, smoking history, age) and time varying factors (e.g., seasonal and diurnal patterns).
Animal Toxicology
Preference is given to studies utilizing experimental and control groups that are matched for individual level characteristics (e.g., strain, sex, body weight, litter size, food and water consumption) and time varying factors (e.g., seasonal and diurnal patterns).
Epidemiology
Factors are considered to be potential confounders if demonstrated in the scientific literature to be related to health effects and correlated with PM. Not accounting for confounders can produce artefactual associations; thus, studies that statistically adjust for multiple factors or control for them in the study design are emphasized. Less weight is placed on studies that adjust for factors that mediate the relationship between PM and health effects, which can bias results toward the null. Confounders vary according to study design, exposure duration, and health effect and may include, but are not limited to the following: Short-term exposure studies: Meteorology, day of week, season, medication use, allergen exposure, and long-term temporal trends. Long-term exposure studies: Socioeconomic status, race, age, medication use, smoking status, stress, noise, and occupational exposures.
Statistical Methodology
Controlled Human Exposure
Statistical methods should be clearly described and appropriate for the study design and research question (e.g., correction for multiple comparisons). Generally, statistical significance is used to evaluate the findings of controlled human exposure studies. However, consistent trends are also informative. Detection of statistical significance is influenced by a variety of factors including, but not limited to, the size of the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion; ideally, the sample size should provide adequate power to detect hypothesized effects (e.g., sample sizes less than 3 are considered less informative). Because statistical tests have limitations, consideration is given to both trends in data and reproducibility of results.

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of particulate matter.

Animal Toxicology
Statistical methods should be clearly described and appropriate for the study design and research question (e.g., correction for multiple comparisons). Generally, statistical significance is used to evaluate the findings of animal toxicology studies. However, consistent trends are also informative. Detection of statistical significance is influenced by a variety of factors including, but not limited to, the size of the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion; ideally, the sample size should provide adequate power to detect hypothesized effects (e.g., sample sizes less than 3 are considered less informative). Because statistical tests have limitations, consideration is given to both trends in data and reproducibility of results.
Epidemiology
Multivariable regression models that include potential confounding factors are emphasized. However, multipollutant models (more than two pollutants) are considered to produce too much uncertainty due to copollutant collinearity to be informative. Models with interaction terms aid in the evaluation of potential confounding as well as effect modification. Sensitivity analyses with alternate specifications for potential confounding inform the stability of findings and aid in judgments of the strength of inference from results. In the case of multiple comparisons, consistency in the pattern of association can increase confidence that associations were not found by chance alone. Statistical methods that are appropriate for the power of the study carry greater weight. For example, categorical analyses with small sample sizes can be prone to bias results toward or away from the null. Statistical tests such as *t*-tests and Chi-squared tests are not considered sensitive enough for adequate inferences regarding PM-health effect associations. For all methods, the effect estimate and precision of the estimate (i.e., width of 95% CI) are important considerations rather than statistical significance.

[a](U.S. EPA, 2008).

[b]Murgia et al. (2014); Weakley et al. (2013); Yang et al. (2011); Heckbert et al. (2004); Barr et al. (2002); Muhajarine et al. (1997); Toren et al. (1993); Burney et al. (1989).

[c]UFPs are defined as particles <100 nm in size, but studies often include size fractions larger than 100 nm in the assessment of the relationship between UFP exposure and health effects.

[d]Many factors evaluated as potential confounders can be effect measure modifiers (e.g., season, comorbid health condition) or mediators of health effects related to PM (comorbid health condition).

the relationship between an air pollutant and health can vary depending on the specific pollutant being assessed.

1.1 References

Barr, RG; Herbstman, J; Speizer, FE; Camargo, CA, Jr. (2002). Validation of self-reported chronic obstructive pulmonary disease in a cohort study of nurses. Am J Epidemiol 155: 965-971. http://dx.doi.org/10.1093/aje/155.10.965

Burney, PG; Laitinen, LA; Perdrizet, S; Huckauf, H; Tattersfield, AE; Chinn, S; Poisson, N; Heeren, A; Britton, JR; Jones, T. (1989). Validity and repeatability of the IUATLD (1984) Bronchial Symptoms Questionnaire: an international comparison. Eur Respir J 2: 940-945.

Heckbert, SR; Kooperberg, C; Safford, MM; Psaty, BM; Hsia, J; McTiernan, A; Gaziano, JM; Frishman, WH; Curb, JD. (2004). Comparison of self-report, hospital discharge codes, and adjudication of cardiovascular events in the Women's Health Initiative. Am J Epidemiol 160: 1152-1158. http://dx.doi.org/10.1093/aje/kwh314

Kilkenny, C; Browne, WJ; Cuthill, IC; Emerson, M; Altman, DG. (2010). Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research [Review]. PLoS Biol 8: e1000412. http://dx.doi.org/10.1371/journal.pbio.1000412

Klimisch, HJ; Andreae, M; Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol 25: 1-5. http://dx.doi.org/10.1006/rtph.1996.1076

Muhajarine, N; Mustard, C; Roos, LL; Young, TK; Gelskey, DE. (1997). Comparison of survey and physician claims data for detecting hypertension. J Clin Epidemiol 50: 711-718. http://dx.doi.org/10.1016/S0895-4356(97)00019-X

Murgia, N; Brisman, J; Claesson, A; Muzi, G; Olin, AC; Torén, K. (2014). Validity of a questionnaire-based diagnosis of chronic obstructive pulmonary disease in a general population-based study. BMC Pulm Med 14: 49. http://dx.doi.org/10.1186/1471-2466-14-49

Rooney, AA; Boyles, AL; Wolfe, MS; Bucher, JR; Thayer, KA. (2014). Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect 122: 711-718. http://dx.doi.org/10.1289/ehp.1307972

Toren, K; Brisman, J; Jarvholm, B. (1993). Asthma and asthma-like symptoms in adults assessed by questionnaires: A literature review [Review]. Chest 104: 600-608. http://dx.doi.org/10.1378/chest.104.2.600

U.S. EPA (U.S. Environmental Protection Agency). (1991). Guidelines for developmental toxicity risk assessment (pp. 1-71). (EPA/600/FR-91/001). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=23162

U.S. EPA (U.S. Environmental Protection Agency). (1996). Guidelines for reproductive toxicity risk assessment (pp. 1-143). (EPA/630/R-96/009). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum.

U.S. EPA (U.S. Environmental Protection Agency). (1998). Guidelines for neurotoxicity risk assessment [EPA Report] (pp. 1-89). (EPA/630/R-95/001F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://www.epa.gov/risk/guidelines-neurotoxicity-risk-assessment

U.S. EPA (U.S. Environmental Protection Agency). (2005). Guidelines for carcinogen risk assessment [EPA Report] (pp. 1-166). (EPA/630/P-03/001F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. http://www2.epa.gov/osa/guidelines-carcinogen-risk-assessment

U.S. EPA (U.S. Environmental Protection Agency). (2008). Integrated science assessment for sulfur oxides: Health criteria [EPA Report]. (EPA/600/R-08/047F). Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment- RTP. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=198843

U.S. EPA (U.S. Environmental Protection Agency). (2013). Toxicological review of trimethylbenzenes (CASRN 25551-13-7, 95-63-6, 526-73-8, and 108-67-8) in support of summary information on the Integrated Risk Information System (IRIS): revised external review draft [EPA Report]. (EPA/635/R13/171a). Washington, D.C.: U.S. Environmental Protection Agency, National Center for Environmental Assessment. http://yosemite.epa.gov/sab/SABPRODUCT.NSF/b5d8a1ce9b07293485257375007012b7/ee1e280e77586de985257b65005d37e7!OpenDocument

U.S. EPA (U.S. Environmental Protection Agency). (2015). Preamble to the integrated science assessments [EPA Report]. (EPA/600/R-15/067). Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, RTP Division. https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=310244

von Elm, E; Altman, DG; Egger, M; Pocock, SJ; Gøtzsche, PC; Vandenbroucke, JP. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies [Review]. PLoS Med 4: e296. http://dx.doi.org/10.1371/journal.pmed.0040296

Weakley, J; Webber, MP; Ye, F; Zeig-Owens, R; Cohen, HW; Hall, CB; Kelly, K; Prezant, DJ. (2013). Agreement between obstructive airways disease diagnoses from self-report questionnaires and medical records. Prev Med 57: 38-42. http://dx.doi.org/10.1016/j.ypmed.2013.04.001

Yang, CL; To, T; Foty, RG; Stieb, DM; Dell, SD. (2011). Verifying a questionnaire diagnosis of asthma in children using health claims data. BMC Pulm Med 11. http://dx.doi.org/10.1186/1471-2466-11-52